

## Regressziószámítás alkalmazása kistérségi adatokon

 Szakálné Kanó Izabella<sup>1</sup>

*A lokális térségek versenyképessége a globalizáció folyamatával párhuzamosan egyre fontosabb kérdéssé válik. Ezt alátámasztja az is, hogy az Európai Unió 2007 és 2013 közötti regionális politikája kiemelten foglalkozik a versenyképességgel, valamint az azt befolyásoló tényezőkkel. Ennek közvetlen következménye, hogy a magyar területfejlesztés eredményességének alapvető feltétele a versenyképességet, még pontosabban a kistérségek versenyképességét meghatározó folyamatok megismerése.*

*A versenyképesség mérhetősége és modellezhetősége a fogalom előtérbe helyeződésével együtt került a vizsgálatok fókuszába. A versenyképességi cylinder, a versenyképességi fa és a piramis-modell azt írják le, hogy a területi versenyben való helytállásnak melyek az alapvető feltételei. Ezen modellek alapján készültek korábban vizsgálatok mind egyszerű adatelemzési, mind többváltozós elemzési módszerek segítségével.*

*A tanulmány célja a kistérségek versenyképességét leíró piramis-modell alapkategóriái és alaptényezői közötti összefüggés elemzése regressziószámítás segítségével. Konkrét összefüggések feltárása a magyar kistérségek megvalósult versenyképessége és az arra ható tényezők között.*

*Kulcsszavak: kistérségek versenyképessége, piramis-modell, regressziószámítás, súlyozott legkisebb négyzetek módszere.*

### 1. Bevezetés

Az Európai Unió bővítésének folyamatának és a globalizációnak hatására olyan fogalmak kerültek előtérbe Magyarországon is, mint a versenyképesség, a területi verseny és az innováció. Világossá vált, hogy a területi tervezés és fejlesztés az egyik lényeges tényezője az ország sikerességének. Az Európai Unió 2007-2013-as regionális támogatásai során kiemelt fontosságot tulajdonítanak a lokális térségek versenyképessége javításának.

Az Európai Unió regionális politikájában évek óta a tartós sikeresség, a fenntartható gazdasági fejlődés a versenyképesség fogalmából indul ki. A versenyképesség fogalma pedig a területi verseny szemléletére támaszkodik. A területi versenyben való sikeresség előfeltételeinek és különböző befolyásoló tényezőinek ismerete célzottá és ez által hatékonyabbá teheti az esetleges beavatkozásokat. Ahhoz, hogy ezeket a fo-

---

<sup>1</sup> Szakálné Kanó Izabella, tanársegéd, Szegedi Tudományegyetem Gazdaságtudományi Kar Közgazdaságtani és Gazdaságfejlesztési Intézet (Szeged)

lyamatokat, összefüggéseket felmérjük, meg kell mondanunk, hogy miben mérjük a szóban forgó terület – *ország, régió* vagy *kistérség*, illetve *város* – versenyképességét. Tanulmányunkat e két fogalom széles körben elfogadott definíciójára építjük.

A *területi verseny* „... egy olyan folyamat, amely a területi egységek között zajlik, és célja a régióban, városban élők jólétének növelése a regionális, helyi gazdaság fejlődésének elősegítésével, amely fejlődést bizonyos csoportok a helyi politikákon keresztül más térségekkel versengve, rivalizálva próbálnak befolyásolni explicit vagy gyakran implicit módon” (Lengyel 2003, 153.o.).

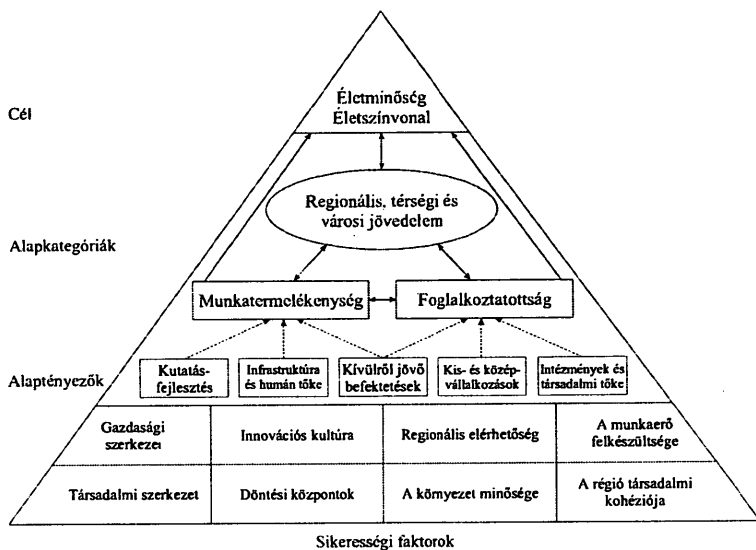
A *versenyképesség* „... a vállalatok, iparágak, régiók, nemzetek és nemzetek feletti régiók képessége *relatív magas jövedelem* és *relatív magas foglalkoztatottsági szint* tartós létrehozására, miközben a nemzetközi (globális) versenynek ki vannak téve” (EC 1999, 75. o., Lengyel 2000, 974. o.). Tehát a versenyképesség két, egymással időnként ellentmondó feltételt fogalmaz meg: a magas foglalkoztatottsági szint melletti tartós gazdasági növekedést.

A versenyképesség fenti definíciója alapján építhető fel a regionális versenyképesség piramis-modellje (1. ábra). A modell a versenyképességgel kapcsolatos területi jellemzőket három nagy csoportra osztja. Így beszélhetünk a versenyképesség mérését lehetővé tevő *alapkategóriákról*, a versenyképességre rövidebb távon, közvetlenül ható *alaptényezőkről*, és a hosszabb távú, közvetett hatással bíró *sikerességi faktorokról*.

Dolgozatomban a piramis-modell kistérségi alkalmazhatóságának elemzését végzem el két szint, az alapkategóriák és az alaptényezők közti kapcsolat számszerűsítésével. Mivel az alaptényezők NUTS 2 szintű régiókra lettek kidolgozva, lehetséges, hogy alacsonyabb aggregáltsági szintre történő alkalmazásakor (pl. kistérség vagy megye) már nincs minden tényezőnek létjogosultsága. Ennek a feltételezésnek is utána kívánok járni, amikor a kapcsolatok szorosságát elemzem a különböző alaptényezők és a megvalósult, mérhető versenyképesség között.

Tanulmányomban a lineáris regressziós alapmodell egyenleteit követően áttekintem az adatok körét, a piramis-modell elemeinél felhasznált egyes mérőszámokat. Ezt követően ismertetem a regressziós modell becsléséhez vezető utat, és az ebből levonható következtetéseket. Ezután bemutatom a végső modellt, amely már megfelel a lineáris regressziós modell alapfeltételezéseinek, majd elemzem és értelmezem a kapott együtthatókat. Tanulmányom utolsó részében áttekintem a további vizsgálatok lehetséges irányait, végül összefoglalom a vizsgálat eredményeit.

## 1. ábra A régiók versenyképességének piramis-modellje



Forrás: Lengyel (2003, 292. o.)

## 2. A regressziós modell, a vizsgált adatok köre és a hipotézisek

Az elemzés során használni kívánt eszköz a *többszörös lineáris regresszió* (Kovács és szerzőtársai 2006, Maddala, G.S. 2004, Ramanathan 2002), amely a legkisebb négyzetek módszerén alapul. Ehhez szükség van egy eredményváltozóra (továbbiakban  $y$ ) és egy vagy több magyarázóváltozóra (a továbbiakban  $x_1, x_2, x_3, \dots$ ). Feltételezésünk szerint az eredményváltozó és a magyarázóváltozók között fennáll egy lineáris összefüggés (konstans tag, plusz a magyarázó változók lineáris kombinációja), az ettől való eltérések pedig egy hibatagba sűrítethetők, amely 0 várható érték valószínűségi változó. A modell tehát a következőképpen írható fel:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon \quad (1)$$

Célunk a  $\beta$  együtthatók becslése ( $\hat{\beta}$ ), amelyek segítségével felírható a becslőfüggvény:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_m x_{im} \quad (2)$$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_m x_{im} + e_i \quad (3)$$

$$e_i = y_i - \hat{y}_i \quad (4)$$

ahol  $e_i$  a maradéktag vagy reziduum.

A modell csak akkor használható teljes körűen, ha megfelel a következő feltétel-rendszernek:

1. *Autokorreláció* nem lép fel, vagyis a hibatagok ( $\varepsilon_i$ ) egymástól függetlenek.
2. *Homoszkedaszticitás* áll fenn, vagyis a hibatagok ( $\varepsilon_i$ ) varianciái azonosak.
3. *Multikollinearitás* nem lép fel, vagyis a magyarázóváltozók között nincs számottevő lineáris kapcsolat.

A korábban már leírtak miatt feltételezésem szerint a piramis-modell alaptényezői és a megvalósult versenyképesség között ok-okozati összefüggés áll fenn. Így eredményváltozóként a megvalósult versenyképesség valamely mérőszámát kívántam használni, a modell magyarázóváltozóinak szerepében pedig az alaptényezők valamilyen (később pontosabban meghatározott) reprezentációit.<sup>2</sup>

Lukovics (2007) a 168 magyar kistérség 2004-es évi alapadataival (279 db) dolgozott, amikor a piramis-modell statisztikai tartalommal való feltöltését végezte. Potenciális indikátorok (138 db) képzése után főkomponens analízis segítségével szelektálta azokat, majd az így megkapott 78 db releváns indikátorral végezte további vizsgálatait. Adatainak forrása a TEIR adatbázisa és a KSH központi adatbázisa (KSH 2005) volt. A mutatók kiválasztása nemzetközileg elismert kutatásokra épült. A 78 mutatóból dolgozatomban 47 db-ot használtam fel (lásd 1. Melléklet), a 168 kistérség 2004-es adatait (Lukovics 2007).

Mivel a tanulmány célja a piramis-modellben szereplő két szint közötti összefüggés feltárása – a regressziószámítás feltételeinek megfelelően –, egyetlen változóba (megvalósult versenyképesség) kell összesűriteni az alapkategóriákban rejlő lehető legtöbb információt. Ennek eszközeként a *főkomponens analízist*<sup>3</sup> (Kovács és szerzőtársai 2006) használtuk. Először is a főkomponensanalízist a 15 változóra lefuttatva azt kaptuk, hogy négy főkomponenshez tartozik egynél nagyobb sajátérték, és az első főkomponens a 15 mutatószám információtartalmának 52%-át őrizte meg. Célunk az volt, hogy ez az érték legalább 70% legyen, így a változók szelektálásával és minden lépésben újabb főkomponensanalízis elvégzésével végül a hetedik lépésben jutottunk el célunkhoz: az első főkomponens 9 változó információtartalmának 71,5%-át őrizte meg.

Az alapkategóriák főkomponensét az alábbi 9 változóból képeztük (az SPSS

<sup>2</sup> E mérőszámok kiválasztásában alapvetően támaszkodtam Lukovics Miklós „*A lokális térségek versenyképességének elemzése*” című (2007) doktori értekezésében, a modell specifikációját illetően pedig Kovács Péter „*A multikollinearitás vizsgálata lineáris regressziós modellekben (a Petres-féle RED-mutató)*” című (2008) doktori értekezésében leírtakra.

<sup>3</sup> A vizsgálat ezen részét Lukovics Miklóssal együtt végeztük.

program használata során ez a főkomponens a későbbiekben Alapkategóriákból képzett főkomponens néven szerepel majd, és versenyképességi mutatóként, illetve eredményváltozóként hivatkozunk rá):

1. Az egy adózóra jutó adóköteles jövedelmek (AEE+ EVA-alap + összevont adóalap, leosztva az adózók adónemenkénti számának összegével).
2. Az egy lakosra jutó személyi jövedelemadó alapot képező jövedelem.
3. Az egy adófizetőre jutó munkaviszonyból származó jövedelem.
4. Egy lakosra jutó bruttó hozzáadott érték.
5. Az egy foglalkoztatottra jutó bruttó hozzáadott érték.
6. Az egy adózóra jutó személyi jövedelemadó alap.
7. A foglalkoztatottsági ráta.
8. A munkanélküliségi ráta.
9. A személyi jövedelemadót fizetők ezer lakosra jutó száma.

A vizsgálat célja az alaptényezők és alapkategóriák közötti kapcsolat számszerűsítése a lineáris regressziós modell segítségével. A vizsgálat eredménye tehát regressziós együtthatók egy csoportja, amelyeket aztán közgazdaságtanilag lehet tovább elemezni. Ezen regressziós együtthatók néhány tulajdonságáról azonban már a vizsgálat elvégzése előtt feltételezésekkel éltem, így 3 hipotézist fogalmaztam meg:

1. *Hipotézis: az alaptényezők mutatószámai mind szignifikáns kapcsolatban állnak az általunk kreált versenyképességi mutatószámmal.* A térségek versenyképességét leíró piramis-modell felépítése alapján ugyanis feltételezzük egy ok-okozati kapcsolatot a modell alaptényezői és alapkategóriái között, vagyis feltételezzük, hogy az alaptényezők fejlesztése vezet a megvalósult versenyképesség magasabb szintjéhez.
2. *Hipotézis: a Kutatás-fejlesztés, technológia, innovációs kapacitás alaptényező hatóköre messze túlmutat a kistérségi határokon, így a regressziós modellben nem kap szerepet.* Mivel az alaptényezők NUTS 2 szintű régiókra lettek kidolgozva, feltételezésem szerint a modell kistérségi szintre történő alkalmazásakor már nincs minden alaptényezőnek létjogosultsága (Lengyel 2003, Lukovics 2007).
3. *Hipotézis: két mutatószám (a Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva és a Nyugdíjban, nyugdíjszerű ellátásban részesülők ezer lakosra jutó száma) kivételével az összes alaptényezőben előforduló indikátor pozitívan korrelál a versenyképességi mutatóval, és – amennyiben bekerülnek a regressziós egyenlet magyarázóváltozói közé – a hozzájuk tartozó regressziós együtthatók is pozitívak lesznek.*  
A közgazdaságtani tartalmuk alapján várhatóan két mutatószámot kivéve: minden mutatószám növekedése (ceteris paribus) a versenyképességi mutatószám növekedésével jár.

### 3. A regressziós modell becsléséhez vezető út

A vizsgálat célja tehát egy lineáris regressziós modell felállítása a versenyképességi mutatószámokkal, mint eredményváltozóval és az alaptevékenységekben megjelenő 32 indikátor közül néhányat, mint magyarázóváltozóval.

A jól értelmezhető és elemezhető lineáris regressziós modell felállításához először megvizsgáljam az eredményváltozó és a magyarázóváltozók közötti lineáris kapcsolat *relációs együhtartókat*. Ezután az SPSS program segítségével lineáris regressziós modelleket állítottam fel. Az SPSS a legkisebb négyzetek módszerét alkalmazza a regressziós együhtartók kiszámítására.

#### 3.1. A 32 magyarázóváltozós modell vizsgálata

##### Korreláció

Először is, elvégezve a korrelációs vizsgálatot, arra az eredményre jutottam, hogy a 32 alaptevékenységi változó *mindegyike szignifikáns kapcsolatot mutat a versenyképességi mutatószámokkal*. Ez teljesen összhangban van feltevéseinkkel, de az alaptevékenységi változók, vagyis a magyarázóváltozóként alkalmazandó mutatók között is hasonlóképpen szignifikáns kapcsolat áll fenn. Ez pedig kétségbe vonja a legkisebb négyzetek módszerének alkalmazhatóságát, illetve az eredmények elemezhetőségét, mert valószínűleg multikollinearitást von maga után.

##### Regresszió

Az SPSS regresszió-számítási programját lefuttattam a 32 alaptevékenységi változóra. Az eljárás során a forward módszert alkalmaztam. A program 7 változót engedett be a modellbe (1. táblázat).

##### Autokorreláció

A Model summary táblázat alapján a Durbin-Watson teszt értéke  $d = 1,871$ , így 5%-os szignifikancia szinten 7 magyarázó változó esetén  $d_u = 1,836 < 1,871$ , nullhipotézisünket elfogadjuk, a hibátagokat nem tekintjük autokorrektáknak.

##### Homoszkedaszticitás

Megvizsgálva a maradátkiagok abszolút értéke  $|e_i|$ , a becslési értékek  $y$  és a magyarázóváltozók közötti kapcsolatot (2. táblázat), azt kapjuk, hogy több is szignifikánsan különbözik a nullától 5%-os szignifikancia szinten (a  $t$  eloszlás táblázata alapján  $r$  korrelációs együhtartók értéke maximum 0,15 lehetne, hogy elfogadjuk a homoszkedaszticitásra irányuló hipotézist), így a modell heteroszkedasztikusnak tekinthető.

1. táblázat A modell magyarázóváltozói, regressziós együtthatók és standard hibájuk

Változó	$\hat{\beta}_i$	Standard hiba
38. A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában	$8,413 \cdot 10^{-2}$	$1,268 \cdot 10^{-2}$
35. Az egyetemet, főiskolát végzett foglalkoztatottak aránya az összes foglalkoztatotton belül	-0,213	$1,977 \cdot 10^{-2}$
29. A kistérség vállalkozásainak ezer lakosra jutó mérleg főösszege	$1,107 \cdot 10^{-4}$	$2,26 \cdot 10^{-5}$
32. Külföldi érdekeltségű vállalkozások saját tőkéjének egy lakosra jutó értéke	$1,554 \cdot 10^{-4}$	$2,58 \cdot 10^{-5}$
37. A 25 évnél idősebb főiskolai, egyetemi diplomával rendelkező népesség aránya a megfelelő korúak %-ában	0,184	0,041
28. A kistérség vállalkozásainak ezer lakosra jutó saját tőke összege	$-1,169 \cdot 10^{-7}$	$3,98 \cdot 10^{-8}$
45. Nyugdíjban, nyugdíjszerű ellátásban részesülők ezer lakosra jutó száma	$-2,643 \cdot 10^{-3}$	$0,900 \cdot 10^{-3}$
Konstans tag	-0,649	0,308

Forrás: saját szerkesztés

A modell az eredményváltozó szórásnégyzetének 90,9%-át magyarázza meg.

### Multikollinearitás

A *VIF* mutatókat az együtthatók táblázatából kiolvassva rendre a következő értékeket kapjuk: 20,759; 13,126; 14,843; 1,640; 44,711; 12,550; 1,662. Némelyik ezek közül elfogadható volna, de soknak az értéke magas, irodalmi adatok szerint maximálisan az 5 körüli értékek fogadhatók el. A modellre kiszámoltuk a *Petres féle Red-mutató* (Kovács és szerzőtársai 2004, 2005, 2006) értékét is. A sajátértékek: 0,01453; 0,040628; 0,084243; 0,57398; 0,7995; 1,3373; 4,1498; ez alapján pedig *Red* = 0,5561. Ez azt jelenti, hogy a magyarázóváltozók adatállományában a hasznos információt tartalmazó adatok aránya az azonos méretű, minimális redundanciájú adatállományhoz képest 44,39%. A mutatók alapján tehát a modellben fellépő *multikollinearitás mértéke nagy*.

Az LNM alkalmazhatóságának három feltételéből csak egy teljesül, ezért ezt a modellt elvettem, ezután más irányból próbáltam közelíteni a problémát.

2. táblázat A reziduumok abszolút értékei és az egyes magyarázóváltozók közti lineáris korrelációs együtthatók

Változók	Az egyes változók $ e $ -vel vett korrelációja
Reziduumok abszolút értéke ( $ e $ )	1
Becsült érték ( $\hat{y}$ )	0,235*
A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában	0,142
A 25 évnél idősebb főiskolai, egyetemi diplomával rendelkező népesség aránya a megfelelő korúak %-ában	0,147
A vállalkozások saját tőke összege 1000 fő	0,262*
Nyugdíjban, nyugdíjszerű ellátásban részesülők ezer lakosra jutó száma	-0,200*
Az egyetemet, főiskolát végzett foglalkoztatottak aránya az összes foglalkoztatotton belül	0,058
A vállalkozások mérlegfőösszege /1000fő	0,245*
Külföldi érdekeltségű vállalkozások saját tőké értéke/lakos	0,032

Forrás: saját szerkesztés

Megjegyzés: \* A korreláció szignifikáns 1 %-os szinten.

### 3.2. Alaptényezőnkénti vizsgálat

Ezután a regressziós vizsgálatot az egyes alaptényezőkre leszűkítve végeztem el. Itt most csak a modellek magyarázóerejét mutatom be (3. táblázat), vagyis azt, hogy az egyes tényezőkben lévő mutatók milyen arányban magyarázzák az eredményváltozó szórásnégyzetét.

3. táblázat Az egyes alaptényezők elemeivel lefuttatott lineáris regressziós modellek magyarázóereje

Tényező	magyarázóerő
II.1. Kutatás-fejlesztés, technológia, innovációs kapacitás	35,5%
II.2. Kis- és középvállalkozások, vállalati szektor	70,9%
II.3. Külföldi működő tőke	35,2%
II.4. Humán tőke és infrastruktúra	84,3%
II.5. Intézmények és társadalmi tőke	49,6%

Forrás: saját szerkesztés



Feltételezésünk – mely szerint nem biztos, hogy minden alaptényezőnek van a piramis-modell középső szintjén létjogosultsága – itt részben igazolódott. Ugyanis a piramis-modellt magasabb aggregáltsági szintre dolgozták ki, és az itt feltüntetett alaptényező sorrend egyben megfelel az Európai Unió fejlesztési prioritási sorrendjének. Ehhez képest – kistérségi szinten – *II.1. A kutatás-fejlesztés, technológia, innovációs kapacitás (K+F)* alaptényező viszonylag alacsony magyarázóerővel bír az általunk meghatározott versenyképességi mutatószám szórásnégyzetében, bár a prioritási sorrend elején szerepel. Ennek oka lehet, hogy a hatása csak magasabb szinten, regionálisan vagy országok szintjén érvényesül. A 32 változóból forward módszerrel kiválasztott magyarázóváltozók között egyetlen mutató sem szerepelt ebből az alaptényezőből. A másik alacsony magyarázóerejű alaptényező (*II.3. Külföldi működő tőke*) változói közül szerepelt a 32. *Külföldi érdekeltségű vállalkozások saját tőkéjének egy lakosra jutó értéke* változó.

### 3.3. A multikollinearitás csökkentése

Sokféle irodalmi adatot (Kovács 2007, Kovács és szerzőtársai 2006, Maddala 2002, Ramanathan 2004, Székelyi–Barna 2003) találtam a multikollinearitás csökkentésének módszereiről, de ezek nem hozták a megfelelő eredményt, vagy közgazdaságtanilag értelmetlenek voltak, vagy pedig alkalmazásuk során más problémák léptek fel. Ezután megpróbáltam olyan modellt alkotni, amely a multikollinearitás problémájával már nem küzd. Ennek szem előtt tartásával szelektáltam a változókat: mivel a *18 évnél idősebb, középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában* változónak a korrelációs együtthatója a legmagasabb a versenyképességi mutatóval, és már önmagában megmagyarázza a versenyképességi mutató szórásának több mint 65%-át, ezt a változót mindenképpen célszerű alkalmazni magyarázóváltozóként.

A multikollinearitás elkerülése végett olyan egyéb mutatószámokat választottam ki a 32 mutató közül, amelyeknek lehetőség szerint alacsony (maximum 0,48) a korrelációs együtthatója a már kiválasztott *18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában* mutatóval. Ezek után már csak 13 változó jöhetett szóba összesen, ezek a láthatók a 4. táblázatban.

A 13 változó tovább csoportosítható az egymás közötti korrelációs együtthatók alapján. Ez többé-kevésbé egybeesik az eredeti alaptényezőkkel. Így a 28. és 30. változók közül és a 31-34., valamint a 44-46. változócsoporthoz is csak egy-egy kerülhetett be a modellbe. Ezek alapján végeztem el a változószelekciót, több modellt is kipróbálva.

A végső modell kiválasztásának szempontjai a következők voltak:

1. A modell magyarázóereje legyen magas, minimum 75%.
2. A modellben minimális legyen a multikollinearitás és az autokorreláció.
3. A modellben minden magyarázóváltozó szignifikáns legyen (így a változószelekcióhoz minden esetben a forward módszert használtam.)

4. Az egyes magyarázóváltozók és az eredményváltozó közötti korrelációs együttható és a regressziós együttható előjele egyezzen meg, közöttük ne legyen ellentmondás.

4. táblázat A multikollinearitás csökkentése érdekében szelektált magyarázóváltozók és a 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában változó közti lineáris korrelációs együtthatók

Változók	Az egyes változók kiemelt mutatóval vett korrelációja
21. K+F beruházások 1000 lakosra	0,353*
30. A vállalkozások jegyzett tőke összege / 1000 fő	0,383*
28. A vállalkozások saját tőke összege / 1000 fő	0,453*
31. Külföldi érdekeltségű váll-ok statisztikai létszáma / 1000 lakos	0,206*
32. Külföldi érdekeltségű vállalkozások saját tőké értéke/lakos	0,368*
33. A külföldi érdekeltségű vállalkozásokban a külföldi tőke összege/lakos	0,373*
34. A külföldi érdekeltségű vállalkozások nettó árbevételének értéke /lakos	0,135
38. A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában	1
42. Az év folyamán kiadott építési engedélyek száma /1000 lakos	0,427*
43. Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva	- 0,286*
46. A működő nonprofit szervezetek ezer lakosra jutó száma	0,479*
44. Belföldi vándorlási különbözet évi átlaga ezer lakosra, 2000-2004	0,270*
45. Nyugdíjban, nyugdíjszerű ellátásban részesülők ezer lakosra jutó száma	- 0,412*

Megjegyzés: \*A korreláció szignifikáns 1 %-os szinten.

Forrás: saját szerkesztés

E négy szempont figyelembevételével a következő változók magyarázzák legjobban az eredményváltozó szórását:

28. A vállalkozások saját tőke összege /1000 fő  
 31. Külföldi érdekeltségű vállalatok statisztikai létszáma 1000 lakosra  
 38. A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában  
 43. Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva

A legkisebb négyzetek módszerének alkalmazása után egy olyan modellt kaptam, amely megfelel a fent leírt szempontoknak, de heteroszkedasztikus. A modell adatai a következők.

### Korreláció

A 4 magyarázóváltozó és az eredményváltozó közötti korrelációs együtthatók az 5. táblázatban láthatók. A versenyképességi mutató és a többi változó közötti kapcsolat szignifikáns már 1%-os szinten is, éppúgy, mint az egymás közötti együtthatók értéke, de azt megállapíthatjuk, hogy a kapcsolat minden változó esetében az eredményváltozóval a legszorosabb.

5. táblázat A szelektált változók közötti lineáris korrelációs együtthatók

Változók	Alapkategóriákból képzett főkomponens	A 18 évnél idősebb közép-fokú végzettséggel ...	Külföldi érdekeltségű váll-ok statisztikai létszáma 1000 lakosra	Korhatár alatti rokkantsági nyugdíjasok aránya ...	A vállalkozások saját tőke összege 1000 fő
Alapkategóriákból képzett főkomponens	1,000	0,810*	0,429*	-0,484*	0,605*
A 18 évnél idősebb közép-fokú végzettséggel ...	0,810*	1,000	0,206*	-0,286*	0,453*
Külföldi érdekeltségű váll-ok statisztikai létszáma 1000 lakosra	0,429*	0,206*	1,000	-0,223*	0,381*
Korhatár alatti rokkantsági nyugdíjasok ...	-0,484*	-0,286*	-0,223*	1,000	-0,273*
A vállalkozások saját tőke összege 1000 fő	0,605*	0,453*	0,381*	-0,273*	1,000

Forrás: saját szerkesztés

Megjegyzés: \*A korreláció szignifikáns 1 %-os szinten.

### Regresszió

Az SPSS a kapcsolat szorosságának sorrendjében engedte be a magyarázóváltozókat a modellbe. Ezek szerint minden együttható szignifikánsan különbözik a 0-tól. A Model Summary táblázat (6. táblázat) alapján látható, hogy a modell magyarázóereje 80,3%.

A modell tehát:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \hat{\beta}_4 x_{i4} \quad (5)$$

ahol

$x_1$ : A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában;

$x_2$ : A vállalkozások saját tőke összege 1000 fő;

$x_3$ : Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva;

$x_4$ : Külföldi érdekeltségű váll-ok statisztikai létszáma 1000 lakosra.

A versenyképességi mutatószámot számszerűsítő statisztikai modell a következő:

$$\hat{y}_i = -1,614 + 0,07266x_{1i} + 0,931 \cdot 10^{-7} x_{2i} - 0,04091x_{3i} + 0,00243x_{4i} \quad (6)$$

6. táblázat A modell magyarázóváltozói, regressziós együtthatók és standard hibájuk

Változó	$\hat{\beta}_i$	Standard hiba
A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában	0,07266	0,00464
A vállalkozások saját tőke összege 1000 fő	$0,931 \cdot 10^{-7}$	$0,196 \cdot 10^{-7}$
Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva;	- 0,04091	0,00718
Külföldi érdekeltségű váll-ok statisztikai létszáma 1000 lakosra	0,00243	0,00052
Konstans	-1,614	0,201

Forrás: saját szerkesztés

7. táblázat Az egyes modellek magyarázóereje és a Durbin-Watson próba eredménye

Modell	R	R <sup>2</sup>	A becslés standard hibája	Durbin-Watson statisztika
1	0,810	0,655	0,589	
2	0,853	0,727	0,525	
3	0,881	0,776	0,478	
4	0,896	0,803	0,450	2,005

Forrás: saját szerkesztés

### Autokorreláció

A 7. táblázat alapján a Durbin-Watson teszt értéke  $d = 2,005$ , 5%-os szignifikancia szinten 4 magyarázóváltozó esetén  $d_U = 1,796$  és  $4 - d_U = 2,204 > d = 2,005$ , nullhipotézisünket elfogadjuk, a hibatagokat nem tekintjük autokorrelálnak.

### Multikollinearitás

A VIF mutatókat a 6. táblázatból kiolvassva rendre a következő értékeket kapjuk: 1,3047; 1,1908; 1,1398; 1,4306. Irodalmi adatok szerint az 5 alatti értékek fogadhatók el, tehát a multikollinearitást nem tekintjük túlzott mértékűnek a modellben.

Még egy multikollinearitási mutatószámot kiszámoltam, ez a Petres-féle RED-mutató. A sajátértékek a következők: 0,49786; 0,76448; 0,81323; 1,9244. Ezek alapján  $Red = 0,3158$ , vagyis a magyarázóváltozók adatállományában a hasznos információt tartalmazó adatok aránya az azonos méretű, minimális redundanciájú adatállományhoz képest 68,42%

### Homoszkedaszticitás

8. táblázat A reziduumok abszolút értékei és az egyes magyarázóváltozók közti lineáris korrelációs együtthatók

Változók	Az egyes változók $ e $ -vel vett korrelációja
Reziduumok abszolút értéke $( e )$	1
Becsült érték $(\hat{y})$	0,261*
A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában	0,089
Külföldi érdekeltségű váll-ok statisztikai létszáma 1000 lakosra	0,331*
Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva	-0,271*
A vállalkozások saját tőke összege 1000 fő	0,316*

Forrás: saját szerkesztés

Megjegyzés: \*A korreláció szignifikáns 1 %-os szinten.

Megvizsgálva a maradéktagok abszolút értéke  $|e_i|$ , a becsült értékek  $\hat{y}_i$  és a magyarázóváltozók közti kapcsolatot (8. táblázat), azt kapjuk, hogy több is szignifikánsan különbözik a nullától 5%-os szignifikancia szinten (a  $t$  eloszlás táblázata alapján  $r$  korrelációs együtthatók értéke maximum 0,15 lehetne, hogy elfogadjuk a homoszkedaszticitásra irányuló hipotézist), így a modellt heteroszkedasztikusnak te-

kintjük. A heteroszkedaszticitás kiküszöbölése érdekében a továbbiakban az úgynevezett *súlyozott legkisebb négyzetek módszerét* alkalmaztam.

#### 4. A végső regressziós modell becslése és hipotézisvizsgálat

A homoszkedaszticitás vizsgálata során kiderült, hogy a hibatagok abszolút értéke szignifikánsan korrelál mind az eredményváltozó becsült értékeinek vektorával, mind a magyarázó változókkal, csak a *18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában* változóval nem. A hibatagok magyarázóváltozóktól való függésének kiküszöbölése érdekében a *súlyozott legkisebb négyzetek módszerét* alkalmaztam. Ennek lényege az, hogy a regresszió során heteroszkedasztikusnak ítélt reziduumok négyzeteiből álló vektornak egy regressziós becslését adjuk meg a heteroszkedaszticitás forrásaként felismert változókkal, azok négyzeteivel és keresztszorzataikkal, mint magyarázóváltozókkal. Ezután az így megbecsült hiba-szórásnégyzetek gyökével (súlyok) elosztva az eredeti egyenletet új regressziós becslést hajtunk végre, ezt már konstans nélkül.

$$\frac{y_i}{w_i} = \hat{\beta}_0 \frac{1}{w_i} + \hat{\beta}_1 \frac{x_{1i}}{w_i} + \hat{\beta}_2 \frac{x_{2i}}{w_i} + \hat{\beta}_3 \frac{x_{3i}}{w_i} + \hat{\beta}_4 \frac{x_{4i}}{w_i} + \frac{e_i}{w_i} \quad (7)$$

Az  $x_i$  változók a már korábban ismertetettek, a  $w_i$  értékek pedig az előbb ismertetett módon kiszámolt súlyok. Az új modell tulajdonságai (9. táblázat):

$$\hat{y}_i = -1,740 + 0,069x_{1i} + 1,022 \cdot 10^{-7} x_{2i} - 0,03408x_{3i} + 0,00577x_{4i} \quad (8)$$

#### Autokorreláció

Az SPSS eredményei alapján a Durbin-Watson teszt értéke  $d = 2,000$ , 5%-os szignifikancia szinten 4 magyarázó változó esetén  $d_U = 1,796$  és  $4 - d_U = 2,204 > d = 2,000$ , nullhipotézisünket elfogadjuk, *a hibatagokat nem tekintjük autokorrelálnak.*

#### Multikollinearitás

A VIF mutatókra az SPSS outputja alapján rendre a következő értékeket kapjuk: 2,169; 2,062; 2,390; 1,577. Irodalmi adatok szerint az 5 alatti értékek fogadhatók el. Kijelenthető tehát, hogy *a multikollinearitás jelen van a modellben, de jelentős gondot nem okoz.*

9. táblázat A súlyozott modell magyarázóváltozói, regressziós együtthatók, és standard hibájuk

Változó	$\hat{\beta}_i$	Standard hiba
A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában	0,069	0,00367
A vállalkozások saját tőke összege 1000 fő	$1,022 \cdot 10^{-7}$	$0,285 \cdot 10^{-7}$
Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva;	- 0,03408	0,00367
Külföldi érdekeltségű váll-ok statisztikai létszáma 1000 lakosra	0,00577	0,00076
Konstans	-1,740	0,154

Forrás: saját szerkesztés

### Homoszkedaszticitás

A súlyozott modellben a korrelációk a maradéktagok abszolút értéke és a magyarázóváltozók, valamint a maradéktagok abszolút értéke és a becült érték között már nem szignifikánsak 5%-os szint alatt, hiszen a 10. táblázatban látható legnagyobb lineáris korrelációs együttható 0,090 és  $\frac{0,090 \cdot \sqrt{1,66}}{\sqrt{1-0,090^2}} = 1,158 < 1,960 = t_{0,05}$ . Tehát a súlyozás valóban kiküszöbölte a maradéktagok eltérő szórásából adódó problémákat.

10. táblázat A súlyozott modellben a reziduumok abszolút értékei és az egyes magyarázóváltozók közti lineáris korrelációs együtthatók

Változók	Az egyes változók korrelációja  e  -vel
Reziduumok abszolút értéke ( e )	1
Becült érték ( $\hat{y}$ / súly)	0,053
A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában / súly	- 0,058
Külföldi érdekeltségű vállalatok statisztikai létszáma 1000 lakosra	0,090
Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva / súly	- 0,027
A vállalkozások saját tőke összege 1000 fő / súly	- 0,055
1 / súly	- 0,052

Forrás: saját szerkesztés

Megjegyzés: \*A korreláció szignifikáns 1 %-os szinten.

A súlyozott legkisebb négyzetek módszerével kapott modell *teljesíti a feltételeket, magyarázóereje kielégítő*, így ezt a modellt fogadtam el végsőként, amit a továbbiakban értelmezni, elemezni kívánok

## 5. Eredmények értelmezése és következtetések

A regressziós elemzés végeredményeként kapott modellt tehát a következő:

$$\hat{y}_i = -1,740 + 0,069x_{1i} + 1,022 \cdot 10^{-7} x_{2i} - 0,03408x_{3i} + 0,00577x_{4i} \quad (8)$$

ahol:

$\hat{y}$  : Alapkategóriákból képzett főkomponens (versenyképességi mutató);

$x_1$ : A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában;

$x_2$ : A vállalkozások saját tőke összege 1000 főre;

$x_3$ : Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva;

$x_4$ : Külföldi érdekeltségű vállalatok statisztikai létszáma 1000 lakosra.

A becslőegyenlet alapján kijelenthetjük, hogy az alapsokaságban, vagyis a magyar kistérségekben:

1. A 18 évnél idősebb középfokú végzettséggel rendelkező népesség arányának a megfelelő korúak csoportjában vett 1 százalékpontos növekedése (csökkenése) esetén (ceteris paribus) a versenyképességi mutatószám átlagosan 0,069 egységgel növekszik (csökken).
2. A vállalkozások saját tőkéje 1000 főre eső összegének 10000 Ft-os növekedése (csökkenése) esetén (ceteris paribus) a versenyképességi mutatószám átlagosan 0,0001022 egységgel növekszik (csökken).
3. A korhatár alatti rokkantsági nyugdíjasok 40-59 éves korosztályhoz viszonyított arányának 1 százalékpontos növekedése (csökkenése) esetén (ceteris paribus) a versenyképességi mutatószám átlagosan 0,034 egységgel csökken (növekszik).
4. A külföldi érdekeltségű vállalatok 1000 lakosra jutó statisztikai létszámának 1 darabbal történő növekedése (csökkenése) esetén (ceteris paribus) a versenyképességi mutatószám átlagosan 0,0058 egységgel növekszik (csökken).

A modellben szereplő regressziós együtthatókról elmondható, hogy előjelük várakozásainknak megfelelően alakultak, vagyis a *Korhatár alatti rokkantsági nyugdíja*-



sok aránya a 40-59 éves korosztályhoz viszonyítva változó regressziós együtthatójának értéke negatív, a többi változó együtthatója pozitív. Ez a mutatók közgazdasági tartalma alapján a 3. számú hipotézist igazolja.

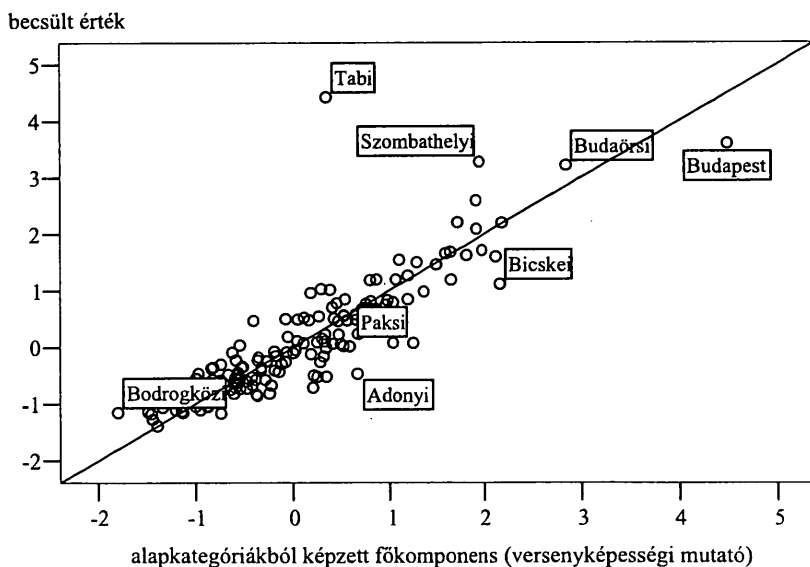
A magyarázóváltozók az 5 versenyképességi alaptényező közül négyből kerültek ki. Ez két említésre méltó következtetés levonását teszi lehetővé.

Egyrészt a 2. hipotézist látszik alátámasztani, amely szerint nem minden alaptényezőnek van létjogosultsága kistérségi szinten a piramis-modellben. A *II.1. Kutatás-fejlesztés, technológia, innovációs kapacitás* alaptényező mutatószámai csak abban az esetben kaptak helyet a magyarázóváltozók között (forward módszer alkalmazása során) amikor a magyarázóváltozó jelöltek kizárólag ebből a változócsoporthoz kerültek ki. A csoport elemeivel, mint magyarázóváltozókkal elvégzett regressziós elemzés modelljének magyarázóereje is alacsony volt: 35,5%.

Másrészt azt láthatjuk, hogy az alaptényezőkön belüli korreláció magas szintje és az alaptényezők közötti korreláció ehhez képest alacsonyabb értéke miatt modellselektációs eljárásunk igazolta, hogy *a versenyképességi mutató az alaptényezők széles talapzatán nyugszik*, hiszen a négy magyarázóváltozó négy különböző alaptényezőtől került ki, köztük fontos és releváns kapcsolat van.

A 2. ábra az eredeti eredményváltozó (alapkategóriákból képzett főkomponens) és annak becslési értékei közötti összefüggést szemlélteti. Az ábrában kiemeltem néhány kistérséget, amelyek a regressziós elemzés szempontjából érdekeseek lehetnek.

2. ábra A versenyképességi mutató és előrejelzése



Forrás: saját szerkesztés

A versenyképességi mutató – amely standardizált változó – értékeit tekintve egy *outlierről* beszélhetünk, ez a *Budapesti kistérség*, hiszen a mutató értéke kívül esik a háromszoros szóráson ( $y_{\{Budapesti\}} = 4,49$ ). Ezen kívül a *Budaörsi kistérség* ( $y_{\{Budaörsi\}} = 2,84$ ) rendelkezik még magas eredményváltozó értékkel, a legalacsonyabb versenyképességi mutató érték pedig a *Bodrogi kistérség* ( $y_{\{Bodrogi\}} = -1,74$ ).

Míg az előbb az eredményváltozó értékeit tekintve különleges eseteket mutatam be, addig érdekes lehet a becslés pontatlansága szempontjából kilógó esetekre is kitérni. Ezek közül is kiemelkedik a *Tabi kistérség*, melynek esetében a modell jóval felülbecsüli a versenyképességi mutatót ( $y_{\{Tabi\}} = 0,35 > 4,42 = \hat{y}_{\{Tabi\}}$ ,  $e_{\{Tabi\}} = 4,07$ ), sőt a Budapesti kistérségnél is versenyképesebbnek mutatja ( $y_{\{Budapesti\}} = 3,60 < 4,42 = \hat{y}_{\{Tabi\}}$ ).

Ez a tény két okból fontos. Egyrészt megmutatja, hogy a súlyozott modell mennyire bünteti a sorból való kilógást. Az eredeti modell ugyanis minden kistérséget azonos súllyal vett be a modellbe. Ennek során a *Tabi kistérség* magas értékű reziduummal rendelkezett, de nem ennyire feltűnő mértékben. A súlyozott modell ezek után épp a reziduumok alapján a *Tabi kistérség* adatait kisebb súllyal szerepeltette, így a többi kistérségre pontosabb modellt kaptunk, mint a korábbi volt. Ez egyben azt is jelenti, hogy a modell a *Tabi kistérség*re nem alkalmazható.

Másrészt a *Tabi kistérség* sorból való kilógásának okait tovább keresve megállapíthatjuk, hogy az egyértelműen a külföldi érdekeltségű vállalkozások kistérségbeli magas számának köszönhető. Ez az eredmény egyébként egyáltalán nem meglepő, hiszen Lukovics Miklós klaszterezési vizsgálatai során is a *Tabi kistérség* különleges viselkedését lehetett megfigyelni (Lukovics 2007).

A többi, az ábrában kiemelt kistérségre vonatkozó reziduumok is nagyok (mind 1-nél nagyobbak), de nem lógnak ki a sorból akkora mértékben, mint a *Tabi kistérség*. A *Szombathelyi kistérség* esetében a regressziós modell felülbecsüli, míg a *Bicskei*, a *Paksi* és az *Adonyi kistérségek* esetében alulbecsüli a versenyképességet:  $e_{\{Szombathelyi\}} = -1,33231$ ;  $e_{\{Bicskei\}} = 1,05165$ ;  $e_{\{Paksi\}} = 1,17147$ ;  $e_{\{Adonyi\}} = 1,13766$ .

Az eredményeim a következő pontokban foglalhatók össze:

1. A 2. hipotézis, mely szerint kistérségi szinten nem minden alaptényezőnek van meghatározó szerepe a megvalósult versenyképesség szintjében, igazolódott, a *Kutatás-fejlesztés, technológia, innovációs kapacitás alaptényező* esetében csak alacsony magyarázóerőt tudtam meghatározni.
2. Az 1. és a 3. hipotézis a korrelációs együtthatók vizsgálatánál már igazolódott, hiszen a versenyképességi mutató és a 32 alapkategóriabeli indikátor között szignifikáns a kapcsolat és a korrelációs együtthatók előjele, illetve a végső regressziós modell együtthatóinak előjele is mind megfelelnek a változók közgazdasági tartalmának.
3. A legmagasabb magyarázóerővel vizsgálataim szerint a *Humán tőke és inf-*

rastruktúra alaptényező rendelkezik, ennek mutatói közül is *A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában* változó mutatta a legmagasabb korrelációt a versenyképességi mutatóval. Ez azt jelenti, hogy az iskolázottság, ezen belül is a érettségi vizsga szintjét elért felnőtt lakosság aránya, vagyis a humán tőke képzettsége meghatározó jelentőséggel bír a versenyképesség elért szintjében. Ennek fejlesztése, több középszintű képzés indítása stb. nagymértékben növelheti a versenyképességet.

4. Mivel az egyes alaptényezőkön belül a mutatók viszonylag homogének voltak, és a négy magasabb magyarázóerővel rendelkező alaptényező már meghatározó jelentőségű a versenyképesség szintjében, ezért végső modellünk mind a négy alaptényezőből egyet-egyet tartalmaz magyarázóváltozóként. Ezek a 11. táblázatban láthatók.

11. táblázat Az egyes alaptényezőkből származó modellben szereplő mutatószámok

Alaptényező	Mutató
II.2. Kis- és középvállalkozások, vállalati szektor	28. <i>A vállalkozások saját tőke összege 1000 fő</i>
II.3. Külföldi működő tőke	31. <i>Külföldi érdekeltségű vállalatok statisztikai létszáma 1000 lakosra</i>
II.4. Humán tőke és infrastruktúra	38. <i>A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában</i>
II.5. Intézmények és társadalmi tőke	43. <i>Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva</i>

Forrás: saját szerkesztés

## 6. A dolgozat korlátai és további lehetséges vizsgálatok

A dolgozat keretei szűkösek, ezért most felhívom a figyelmet néhány további lehetséges elemzés lehetőségére. Vizsgálataim célja egyértelműen az volt, hogy a piramismodellben szereplő, akár gazdasági típusúnak is nevezhető adatok szintjei között találjak összefüggést. Ez azonban nem zárja ki más mutatók, például földrajzi, természeti adottságok számszerűsített változóinak befolyásoló hatását. Sőt, nagy valószínűséggel ezek az adottságok mind az alaptényezők, mind az alapkategóriák változóinak értékeit nagymértékben befolyásolják.

Az elemzés igazolta, hogy *II.1. Kutatás-fejlesztés, technológia, innovációs kapacitás* alaptényező mutatószámai kistérségi szinten nem jelentenek releváns befolyásoló tényezőt a versenyképességi mutatóra nézve. Ennek oka elsősorban az, hogy a

*K+F helyek* nagymértékben *koncentrálódnak*, összevontan működnek, leginkább regionális szerepük van, sok kistérségben nincsenek ilyen típusú centrumok, ezért a kistérségi adataink között sok a nulla. A *K+F* hatásának vizsgálatára így más utat kell keresni. Ennek több módja lehetséges, például:

1. Csak a pozitív adatértékekkel rendelkező kistérségeket vonni be a vizsgálatba (ez a vizsgálatnak a nagyvárosokra való szűkítésével lenne közel egyenértékű).
2. Lehetséges az adatokat regionális, vagy megyei szinten vizsgálni, vagy
3. szomszédossági-agglomerációs súlyozással történhetne a vizsgálat. Tehát a *K+F* helyek adatait simítjuk távolsági vagy szomszédossági adatok alapján (vagyis területi mozgóátlagot számítunk), így a nullák száma csökken. Ebben az esetben az összes kistérség szerepelne a vizsgálatokban.

Tovább lehet vizsgálni az adatokat a piramis-modell két alsó szintje, vagyis a *si-kerességi faktorok* és az alaptényezők közötti kapcsolat feltárása érdekében.

Mivel a vizsgálat célja az volt, hogy az összefüggéseket ok-okozati szempontból elemezze, ezért az *alaptényezők adatainak inkább korábbi évekből kellene származniuk, mint az alapkategóriáknak*, hogy időbeli hatásukat megállapíthassuk. Az alaptényezők mutatószámainak korábbi és a felhasznált 2004-es sorai között nagy valószínűséggel magas a korreláció, és ilyen bőséggel csak a 2004-es adataink álltak rendelkezésre, ezért ettől a problémától eltekintettünk, azzal együtt, hogy a későbbiekben időben több évvel csúsztatott modellel is el kellene végezni a vizsgálatot. Arra is lehetne keresni a választ, hogy milyen időbeli csúsztatás felel meg leginkább, melyiknek nagyobb a magyarázóereje.

Mivel a Budapesti volt az egyetlen olyan kistérség, amelynek mutatószámai rendkívüli módon *outlier* esetnek számítanak, az elemzés további folytatására adna lehetőséget, ha a Budapesti kihagyásával a 167 kistérség adataiból vezetnénk le következtetéseinket. Ezt tetézi az is, hogy az intézményrendszer egy része (kormányzat, oktatás stb.) nagymértékben összpontosul ebben a kistérségben, statisztikai adatai csak itt jelennek meg, holott a többi kistérség érdekeit is szolgálják, ez pedig torzító hatással jár. Ez ugyanaz a probléma, mint a *K+F* helyekkel kapcsolatban lépett fel, megoldása is hasonló lehet (Lukovics 2007, 200. o.).

## 7. Összefoglalás

Az adatok statisztikai elemzésének célja az volt, hogy a *piramis-modell két szintje* – ezek az *alapkategóriák és az alaptényezők* – egymásra épülésének mértékét, módját és irányát meghatározza annak érdekében, hogy az *alaptényezők fejlesztésének* versenyképességre gyakorolt hatását számszerűsíteni lehessen. A vizsgálat során keresztmetszeti adatokkal dolgoztunk, amelyek a 2004-es évre vonatkoznak, és a teljes alapsoka-

ságról (minden kistérségből) rendelkezésre állt adat.

Regressziószámítást alkalmaztam, melynek eredményváltozója egy, az alapkategóriákból képzett főkomponens volt. Elvégeztem az elemzést az összes alapkategóriabeli mutatóval, mint magyarázóváltozóval, majd az egyes alaptényezőkre leszűkítve, utóbbi esetben ezek magyarázóerejét vizsgáltam. Az adatok között nagymértékű lineáris összefüggés volt tapasztalható, ez pedig a modell multikollinearitásához vezetett. Ezért a végső modellt úgy állítottam össze, hogy ennek elfogadható szintre csökkentése mellett közgazdaságilag értelmes legyen, és elegendően magas magyarázóerővel rendelkezzen. Ehhez végül a súlyozott legkisebb négyzetek módszerét kellett alkalmaznom.

A végső modellben legmagasabb magyarázóerővel a 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában indikátor rendelkezett. Ez azt jelenti, hogy az iskolázottság, ezen belül is a érettségi vizsga szintjét elért felnőtt lakosság aránya, vagyis a humán tőke képzettsége meghatározó jelentőséggel bír a versenyképesség elért szintjében.

Az elemzés igazolta, hogy II.1. Kutatás-fejlesztés, technológia, innovációs kapacitás alaptényező mutatószámai kistérségi szinten nem jelentenek releváns befolyásoló tényezőt a versenyképességi mutatóra nézve. Ennek oka pedig az, hogy a  $K+F$  helyek nagymértékben koncentrálnak, összevontan működnek a városokban.

Felhívom a figyelmet még az úgynevezett Budapest-hatásra is: az intézményrendszer egy része nagymértékben összpontosul Budapesten, bizonyos nagyvállalatok is csak itt vannak bejegyezve, statisztikai adataik csak itt jelennek meg, holott a többi kistérség érdekeit is szolgálják, ez pedig a vizsgálatra nézve torzító hatással jár.

#### *Felhasznált irodalom*

- EC 1999: *Sixth Periodic Report on the Social and Economic Situation and Development of Regions in the European Union*. European Commission, Luxembourg.
- Kovács P. – Petres T. – Tóth L. 2004: Adatállományok redundanciájának mérése. *Statisztikai Szemle*, 6-7, 595-604. o.
- Kovács P. – Petres T. – Tóth L. 2005: A New Measure of Multicollinearity in Linear Regression Models. *International Statistical Review (ISR)*, 3, 405-412. o.
- Kovács P. – Petres T. – Tóth L. 2006: *Válogatott fejezetek statisztikából*. JATEPress, Szeged.
- Kovács P. 2008: A multikollinearitás vizsgálata lineáris regressziós modellekben (a Petres-féle RED-mutató). *Doktori értekezés*. SZTE Közgazdaságtudományi Doktori Iskola, Szeged.
- KSH 2005: *Területi Statisztikai Évkönyv 2004*. Központi Statisztikai Hivatal, Budapest.
- Lengyel I. 2000: A regionális versenyképességről. *Közgazdasági Szemle*, 12, 962-987. o.

- Lengyel I. 2003: *Verseny és területi fejlődés: térségek versenyképessége Magyarországon*. JATEPress, Szeged.
- Lukovics M. 2007: A lokális térségek versenyképességének elemzése. *Doktori értekezés*. SZTE Közgazdaságtudományi Doktori Iskola, Szeged.
- Maddala, G. S. 2004: *Bevezetés az ökonometriába*. Nemzeti Tankönyvkiadó, Budapest.
- Ramanathan, R. 2002: *Bevezetés az ökonometriába, alkalmazásokkal*. Panem Kiadó, Budapest.
- Székelyi M. – Barna I. 2003: *Túlélőkészlet az SPSS-hez, Többváltozós elemzési technikákról társadalomkutatók számára*. Typotex Kiadó, Budapest.

*1. Melléklet:***I. Alapkategóriák****I.1. Jövedelmek**

1. Az egy adózóra jutó adóköteles jövedelmek
2. Az egy lakosra jutó személyi jövedelemadó alapot képező jövedelem
3. Az egy adófizetőre jutó munkaviszonyból származó jövedelem
4. Az egy adófizetőre jutó társas vállalkozásból származó jövedelem
5. Egy lakosra jutó bruttó hozzáadott érték

**I.2. Munkatermelékenység**

6. Az egy foglalkoztatottra jutó AEE
7. Az egy foglalkoztatottra jutó bruttó hozzáadott érték
8. Az egy adózóra jutó személyi jövedelemadó alap

**I.3. Foglalkoztatottság**

9. A foglalkoztatottsági ráta
10. A munkanélküliségi ráta
11. A személyi jövedelemadót fizetők ezer lakosra jutó száma

**I.4. Globális integráltság (nyitottság)**

12. Az egy lakosra jutó exportértékesítés nettó árbevétele
13. Az export aránya a bruttó hozzáadott értékből
14. A külföldiek által eltöltött vendégejszakák ezer lakosra jutó száma a kereskedelmi szálláshelyeken
15. A belföldiek által eltöltött vendégejszakák ezer lakosra jutó száma a kereskedelmi szálláshelyeken

**II. Alaptényezők****II.1. Kutatás-fejlesztés, technológia, innovációs kapacitás**

16. A 10000 lakosra jutó szabadalmak évi átlagos száma 2000-2004
17. A 10000 lakosra jutó MTA köztestületi tagok száma
18. A 100000 lakosra jutó K+F helyek száma
19. Az 1000 lakosra jutó K+F helyek tudományos kutatóinak tényleges létszáma
20. Az 1000 lakosra jutó K+F költségek
21. Az 1000 lakosra jutó K+F ráfordítások
22. Az 1000 lakosra jutó K+F beruházások értéke

**II.2. Kis- és középvállalkozások, vállalati szektor**

23. Működő társas vállalkozások ezer lakosra jutó száma
24. Működő társas kisvállalkozások (10-49 alkalmazott) ezer lakosra jutó száma
25. Működő jogi személyiségű vállalkozások ezer lakosra jutó száma
26. Működő jogi személyiségű kisvállalkozások (10-49 alkalmazott) ezer lakosra jutó száma
27. Működő jogi személyiségű vállalkozások aránya a működő gazdasági szervezetekből

28. A kistérség vállalkozásainak ezer lakosra jutó saját tőke összege
29. A kistérség vállalkozásainak ezer lakosra jutó mérleg főösszege
30. A kistérség vállalkozásainak ezer lakosra jutó jegyzett tőke összege

### **II.3. Külföldi működő tőke**

31. Külföldi érdekeltségű vállalkozások statisztikai létszámának 1000 lakosra jutó értéke
32. Külföldi érdekeltségű vállalkozások saját tőkéjének egy lakosra jutó értéke
33. Az 1 lakosra jutó külföldi tőke összege a külföldi érdekeltségű vállalkozásokban
34. A külföldi érdekeltségű vállalkozások nettó árbevételének 1 lakosra jutó értéke

### **II.4. Humán tőke és infrastruktúra**

35. Az egyetemet, főiskolát végzett foglalkoztatottak aránya az összes foglalkoztatotton belül
36. A vezető, értelmiségi foglalkozású foglalkoztatottak aránya az összes foglalkoztatotton belül
37. A 25 évnél idősebb főiskolai, egyetemi diplomával rendelkező népesség aránya a megfelelő korúak %-ában
38. A 18 évnél idősebb középfokú végzettséggel rendelkező népesség aránya a megfelelő korúak %-ában
39. Távbeszélő fővonalak ezer lakosra jutó száma
40. ISDN vonalak 1000 lakosra jutó száma
41. Az év folyamán épített lakások összes alapterülete
42. Az év folyamán kiadott lakásépítési engedélyek 1000 lakosra jutó száma

### **II.5. Intézmények és társadalmi tőke**

43. Korhatár alatti rokkantsági nyugdíjasok aránya a 40-59 éves korosztályhoz viszonyítva
44. Az ezer lakosra jutó belföldi vándorlási különbözet évi átlaga (2000-2004)
45. Nyugdíjban, nyugdíjszerű ellátásban részesülők ezer lakosra jutó száma
46. A működő nonprofit szervezetek ezer lakosra jutó száma
47. A felsőfokú intézményekben nappali tagozatos hallgatók ezer lakosra jutó száma